Toward a Public Dataset of Wide-Field Astronomical Images Captured with Smartphones

Olivier Parisot[a] and Diogo Ramalho Fernandes[b]

Luxembourg Institute of Science and Technology (LIST),
5 Avenue des Hauts-Fourneaux,
4362 Esch-sur-Alzette, Luxembourg

Keywords: Smartphone, Astronomy, Deep Learning.

Abstract: Smartphone photography has improved considerably over the years, particularly in low-light conditions. Thanks to better sensors, advanced noise reduction and night modes, smartphones can now capture detailed wide-field astronomical images, from the Moon to deep sky objects. These innovations have transformed them into pocket telescopes, making astrophotography more accessible than ever. In this paper, we present AstroSmartphoneDataset, a dataset of wide-field astronomical images collected since 2021 with various smartphones, and we show how we have used these images to train Deep Learning models for trails detection.

1 INTRODUCTION

In recent years, astrophotography has become increasingly accessible, whereas it was once limited to experienced enthusiasts with specialized equipment (Woodhouse, 2015). Thus, there are many ways to take pictures of celestial objects such as the Sun, the Moon, the planets of the Solar System and Deep Sky Objects (DSO) – for both amateur and professional astronomers.

Installations combining small optical instruments (such as 60mm diameter refractors) with inexpensive CMOS cameras enable astrophotography and Electronically Assisted Astronomy (EAA, sometimes called *Video Astronomy*) to image most DSO with good details (Ashley, 2017), but the process requires specific equipment and advanced technical skill to set it up and use it properly. To obtain high-quality results, large-diameter instruments coupled with ultra-sensitive CMOS cameras enable displaying stunning images of Jupiter, exploring outer Solar System (Arimatsu, 2025), discovering unknown DSO (Drechsler et al., 2023), or even detecting transient objects (Romanov, 2022). Here again, more expensive and complex equipment is required, as well as considerable know-how.

Recently, modern smart telescopes have taken a major step forward by automating all the tedious tasks involved, making it possible for the general public to observe the sky in a comfortable way (Parisot et al., 2022). As well as being fun and educational, these automated instruments enable astronomers to capture large quantities of data, which is essential for producing quality results (in particular to optimise the signal-to-noise ratio of the images obtained).

In recent years, smartphones have achieved remarkable progress in photography (Fang et al., 2023), both in sensor technology and onboard software processing, matching or even exceeding the performance of digital cameras. It is therefore possible to use smartphones as imagers, whether or not in combination with complementary devices. One example is Hestia, an instrument that amplifies the capabilities of a smartphone to photograph the Moon, the Sun and even the brightest DSO [1].

In this work, we present AstroSmartphoneDataset, a dataset of images captured between 2021 and 2025 with four different smartphones, using the device's built-in sensor and processing, without any additional optical instruments. Such dataset has both academic and practical applications. Scientifically, it enables to design automated systems for satellite detection and space debris monitoring (crucial for Space Situational Awareness), by serving as valuable training data for Deep Learning models in automated object recogni-

[a] https://orcid.org/0000-0002-3293-3628
[b] https://orcid.org/0009-0008-3187-3468

[1] https://vaonis.com/fr/pages/product/hestia

Parisot, O., Fernandes and D. R.
Toward a Public Dataset of Wide-Field Astronomical Images Captured with Smartphones.
DOI: 10.5220/0013642300003967
In *Proceedings of the 14th International Conference on Data Science, Technology and Applications (DATA 2025)*, pages 644-650
ISBN: 978-989-758-758-0; ISSN: 2184-285X

tion and classification. Moreover, it can be used as educational resources for teaching astronomy and image processing techniques.

The present paper is structured as follows. In Section 2, we briefly list recent work related to the use of these devices in astronomy. We present in detail the dataset collected and prepared in Section 3. Finally, we discuss a potential application with streaks detection in Section 4, we list additional potential use-cases in Section 5, and we propose some perspectives in Section 6.

2 RELATED WORKS

Smartphones capable of producing high-quality astronomical images are fairly recent [2], but we can found references such as (Kim and Ju, 2014) or (Sohib et al., 2018) in which the authors combine them and optical devices to enable the practice of EAA. Smartphone manufacturers have put so much effort into their low-light photography features that these objects have begun to be of real interest for capturing night sky images when plugged with other instruments (Falkner, 2020).

Last years, the use of these devices has begun to spread for educational purposes on the subject of capturing wide-field views of constellations from the Southern Hemisphere (Paula and Micha, 2021), for outreach activities to teach astronomy concepts through Moon phases observation (Malchenko, 2024). In (Fisher and Hinckley, 2022), the authors show how modern premium smartphones can be compared to Digital Single-Lens Reflex (DSLR) to take astronomical pictures (for instance of Milky Way). Even NASA has published a detailed guideline to learn how to use smartphones to observe any kind of celestial targets [3]. Nowadays, these devices not only have sensors but also the on-board computing capacity to produce, alone or with telescopes support, formidable results in lunar, solar and deep-sky observation (Lécureuil, 2024).

To the best of our knowledge, there is currently no amateur-accessible database of astronomical images taken with smartphones, such as a sky survey. In the next section, we describe a dataset of such images that we have been collecting and preparing since 2021.

[2] https://en.wikipedia.org/wiki/Camera_phone

[3] https://science.nasa.gov/learn/heat/resource/a-guide-to-smartphone-astrophotography/

3 DATASET DESCRIPTION

AstroSmartphoneDataset, is a collection of wide-field astronomical images captured with Android smartphones by using Pixel Camera (also called Google Camera or GCam), a Google mobile application allowing to take long-exposure astronomical images with a large field of view [4]. The *astrophotography mode* of this mobile application allows to capture and produce long-exposure frames to reduce noise, enhance details, improve dynamic range, while preserving non-astronomical elements like buildings and landscapes, ensuring a natural look.

In order to test Deep Learning algorithms as part of research projects carried out at the Luxembourg Institute of Science and Technology (LIST), data have been collected between 2021 and 2025 from different locations in France. To do this, we have used four smartphones, each with different sensors and onboard calculation capabilities:

- Pixel 4a: 12.2 megapixels rear camera, f/1.7 aperture, HDR+ technology.
- Pixel 6: dual rear cameras, 50 megapixels wide-angle and 12 megapixels ultra-wide.
- Pixel 8a: dual rear cameras, 64 megapixels wide-angle and 13 megapixels ultra-wide.
- Pixel 8Pro: triple rear cameras, 50 megapixels wide-angle, 48 megapixels ultra-wide, and 48 megapixels telephoto.

Data was acquired over a long period of almost four years, in different regions of France (mainly in the north-east in a non-urban area, in the north in the countryside, in the south by the sea, in the centre in the mountains): these different locations have a Bortle level between 4 and 8 (it is a 1-to-9 rating system indicating light pollution levels: the lower the number, the darker the sky and the better the conditions for stargazing (Bortle, 2001)).

As these smartphones were not released at the same time and we did not have them at our disposal when we started collecting data, we used them in chronological order: first the Pixel4a, then the Pixel8Pro, then the Pixel6 and the Pixel8a. The first image was captured on 30 July 2021 with a Pixel4a (Figure 1) while the last image was taken from 18 March 2025 with a Pixel8a.

Data acquisition required the use of often ingenious techniques to orientate and stabilise the phones during capture. Indeed, the idea was to never have exactly the same point of view, so the smartphone used was sometimes placed on a chair, on a table, on the

[4] https://en.wikipedia.org/wiki/Pixel_Camera

Figure 1: First image of AstroSmartphoneDataset, captured with a Pixel4a from the North-East of France in a suburban area (30 July 2021).

floor, held against another object such as a bottle, etc. The main constraint was to ensure that the phone was stable during image capture (requiring the use of the timer, to avoid shaking after pressing the buttons).

Images were taken under different weather conditions (clear sky, foggy sky, almost overcast sky), in order to obtain an heterogeneous set. In practice, the photo mode used means that the stars can be captured on the uncovered parts of the night sky, even if the other part is cloudy.

There was no strict curation of collected images: only really bad images were ignored after capture and then not included in the dataset (for example, blurry images due to incorrect use of the smartphones).

As a result, AstroSmartphoneDataset contains 4635 medium-resolution and 197 high-resolution RGB images in JPEG format (minimal compression), and each RGB image has a minimal resolution of 1440×1920 pixels: they were obtained after registration, stacking and post processing of one-second shots, giving final images corresponding to a cumulative exposure time of 4 minutes.

The GCam application produces two kinds of data: high-resolution images and timelapses (under the form of short MP4 movies) of 1 second. These timelapses are also the result of post-processing (i.e. alignment, stacking and HDR), and we use them to extract additional images, albeit at a lower resolution than the previous images.

The unique characteristic of the images in the dataset is attributable to two factors. Firstly, the wide field of view through which the images were captured, and secondly, the special processing carried out by the GCam application. The resulting images thus combine elements of astronomical images, characterised by a field of stars, and potentially very bright galaxies and nebulae, with elements from the landscape at the time the shot was taken, such as houses, walls, roofs and trees.

AstroSmartphoneDataset is available as an open ZIP archive on Zenodo [5] (Parisot, 2025).

From a purely entertaining point of view, these wide-field images allow you to admire the Milky Way (our galaxy) at different times of the year, as well as other parts of the sky, including well-known constellations such as Orion and the famous Big Dipper. However, we will see in the next few sections that this data can also be used for other purposes.

4 USE CASE: STREAKS DETECTION

A notable feature of the AstroSmartphoneDataset is the significant presence of streaks in the high-resolution images, and particularly in the medium-resolution images derived from the timelapses movies. The aforementioned streaks are typically observed in the form of lines exhibiting varying degrees of thickness and continuity, and are attributed to various sources such as satellites, airplanes, and even meteors (Nir et al., 2018): these lines are often problematic in astrophotography because they lead to data loss and generate additional work to remove them (unpainting).

Traditionally, the detection of such moving objects is a key challenge in Computer Vision, and common techniques include: Background Subtraction (identifies changes between frames by modeling the background), Optical Flow (Estimates pixel motion across frames (Farnebäck, 2003)), Deep Learning (Convolutional Neural Networks or Recurrent Neural Networks for feature extraction and tracking). In (Petit et al., 2004), the authors discuss a method for detecting moving objects in astronomical images using photometric variation analysis and image subtraction techniques: it focuses on improving the detection of faint sources by optimizing background mod-

[5] https://zenodo.org/records/14933725

eling and noise reduction. (Lee, 2023) presents a Deep Learning approach for detecting moving trans-Neptunian objects (TNOs) in large sky surveys: it uses Convolutional Neural Networks trained on artificially generated sources in time-series images from the CFHT MegaCam. The method presented in (Huang et al., 2024) is applied to tracking dynamic space targets, such as satellites and space debris, using image sequences for motion detection and Deep-SORT for tracking. In (Parisot and Jaziri, 2024), the authors combined classification with ResNet50 and GradCAM to detect linear features in raw astronomical images.

A potential application of the AstroSmartphone-Dataset is the construction of an annotated dataset, which can be utilised to train an automatic detection model based on YOLO (You Only Look Once) (Jegham et al., 2025). YOLO is a real-time object detection method that analyze an entire image in a single pass, identifying and localizing multiple objects simultaneously with high speed and accuracy.

As a first step, we have manually selected images, we have split them into 640 × 640 patches and we have annotated the streaks with the MakeSense tool [6], and we have stored images and labels by using the YOLO format: 403 images for training, 88 images for validation, 163 images for test (Figure 2). Given that our images contain night sky, landscape features and streaks to be detected, we made sure to keep background images (i.e. without labels) in each of these sets, to show that features that might look like streaks (house edges, tree branches, etc.) are not.

Then we have trained different versions of YOLO released between 2022 and 2024 (v7, v8 and v11), with different sizes (tiny, normal, extra-large), by using the YOLOv7 official Python package [7] and the Ultralytics Python package [8] without specific customization.

YOLO-v7 implementation uses default hyperparameters such as $lr0 = 0.01$, $momentum = 0.937$, $weight_decay = 0.0005$, with augmentations like $mosaic = 1.0$, $fliplr = 0.5$, and $scale = 0.5$. YOLO-v8, from Ultralytics, follows similar defaults but with slight adjustments and improved augmentation handling. YOLO-v11 extends YOLO-v8 and includes automatic tuning over hyperparameter ranges to optimize training dynamically.

For each type of model, we ran the training in the same way over several hundred epochs (maximum 500), and in most cases, the best model was found after a calculation between 150 or 300 epochs.

[6] https://www.makesense.ai
[7] https://github.com/WongKinYiu/yolov7
[8] https://pypi.org/project/ultralytics/

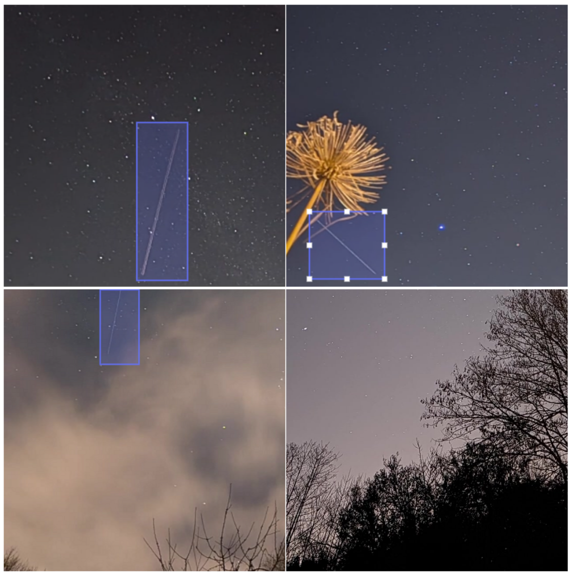

Figure 2: Four 640 × 640 annotated images in the training set: the last one is for background, i.e. without label, to learn the model to ignore undesired patterns in the image.

Based on our tests (Table 1), we found that the best models in terms of accuracy are those with the largest number of parameters, namely yolo-v11-x ($\tilde{5}7$M parameters) and yolo-v7-x ($\tilde{7}1$M parameters). With 26M parameters, yolo-v8-m completes the podium of the best models evaluated, according to the test set.

Figure 4 illustrates the predictions of two models (YOLO-v8-x and YOLO-v11-x) on a sample image from AstroSmartphoneDataset.

This type of model can be used to directly detect whether streaks appear during observations, for monitoring purposes (Space Domain Awareness) or simply to filter/reject impacted images if they should not contain such artefacts. It's a hot topic: the Centre for the Protection of the Dark and Quiet Sky from Satellite Constellation Interference (CPS), established by the International Astronomical Union (IAU) in 2022, coordinates global efforts to mitigate the impact of satellite constellations on astronomy and the night sky.

5 DISCUSSION

As is, AstroSmartphoneDataset contains only images, without labels or metadata, and may be useful for unsupervised or self-supervised learning. It enables tasks like autoencoding, clustering, contrastive learning, and Generative Adversarial Networks (GAN). It can also be leveraged for transfer learning, using pre-trained models (ResNet) to extract features or generate pseudo-labels with AI models like Contrastive Language–Image Pretraining (CLIP) (Radford et al.,

Table 1: Evaluation of the accuracies of the different trained YOLO models – computation was done on the test set (163 images). For each model, the count of parameters is also specified.

Model	Parameters	Precision	Recall	mAP@.5	mAP@.95
yolo-v7	37M	0.553	0.514	0.444	0.187
yolo-v7-tiny	6M	0.575	0.495	0.415	0.146
yolo-v7-x	71M	0.588	0.654	0.568	0.247
yolo-v8-n	3M	0.533	0.495	0.43	0.176
yolo-v8-m	26M	0.474	0.488	0.491	0.224
yolo-v8-x	68M	0.422	0.514	0.412	0.159
yolo-v11-n	2.5M	0.448	0.637	0.492	0.229
yolo-v11-m	20M	0.399	0.486	0.364	0.151
yolo-v11-x	57M	0.704	0.626	0.646	0.332

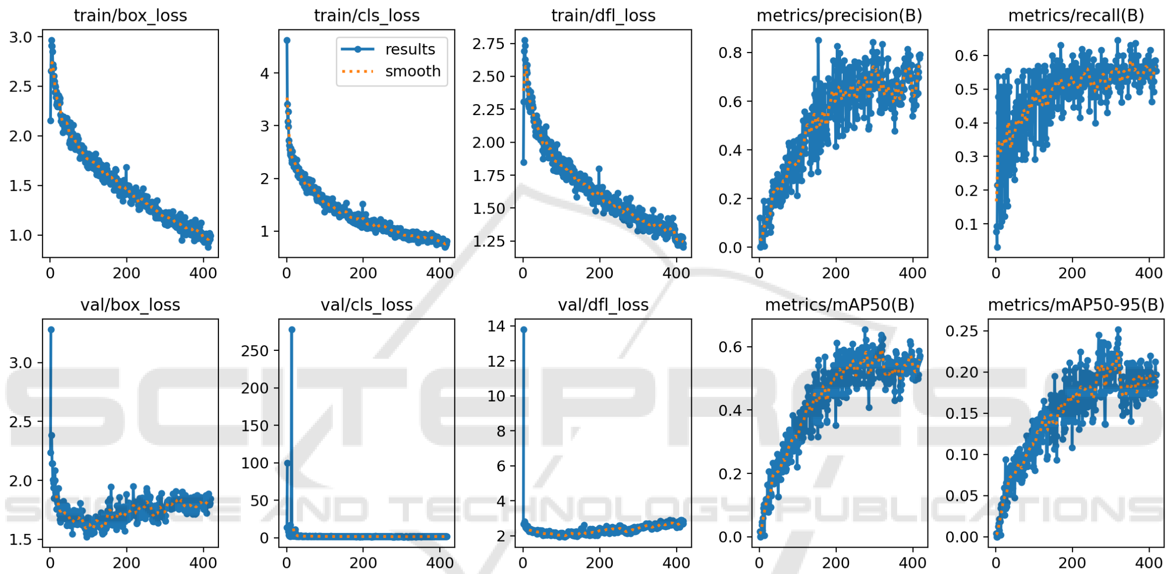

Figure 3: Statistics obtained after the training during 400 epochs of the YOLO-v8-m model with the default hyper-parameters of the Ultralytics Python package.

2021).

Other use cases can be carried out on AstroSmartphoneDataset, with an additional annotation effort (as for the detection of streaks presented in the previous section):

- One example is the possibility of creating an image segmentation model, to be able to clearly distinguish what is sky and what is landscape: YOLO can also be used for this type of task (Kang and Kim, 2023).
- Another example is the design of an Image Quality Assessment (IQA) system for this kind of astronomical images captured with smartphones (IQA), again taking into account the hybrid aspect of images: we could imagine a No Reference IQA system that assesses both the quality of the sky (shape of the stars, darkness of the background, etc.) and the quality of the foreground or background elements: a technique like CLIP-IQA could be a good basis (Wang et al., 2023).

Another potential application of AstroSmartphoneDataset is research about light pollution. By collecting images from locations all over the world, users can contribute to a global analysis of how artificial light impacts visibility in the night sky (Muñoz-Gil et al., 2022). This would assist citizen science projects and environmental studies of light pollution. The large number of images captured, can potentially enable scientists and environmental groups to better understand the patterns and extent of light pollution, a task that would be difficult with stationary and traditional equipment. This would allow amateur astronomers and citizen scientists to contribute together to such global initiatives.

All these perspectives open the door to the creation of new processing methods for the data pro-

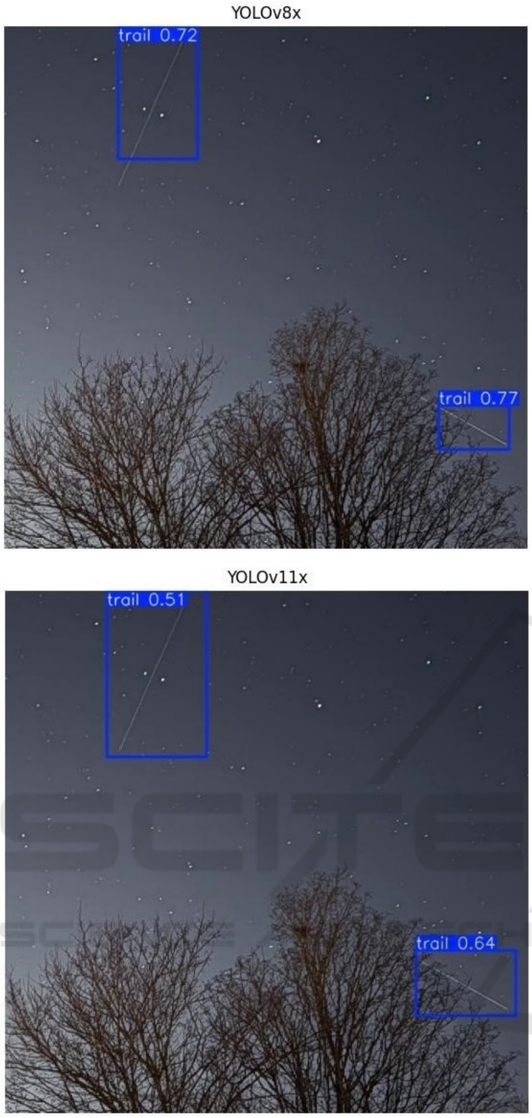

Figure 4: Example of inferences on 640 × 640 images with the trained yolo-8-x and yolo-11-x models: the second one is more efficient to find correctly the bounding boxes around the streaks.

duced by a mechanism similar to that used to generate the images in this dataset.

There is one particular limitation, specific to datasets for Machine Learning/Deep Learning purposes: the models trained/validated on AstroSmartphoneDataset will only be valid and effective if they are applied to similar images, i.e. captured with the same types of smartphone, or at least with similar characteristics (sensors, etc.), but also with similar shooting conditions.

6 CONCLUSION AND PERSPECTIVES

In this paper, we have introduced and described AstroSmartphoneDataset, a collection of wide-field astronomical images captured with various smartphones since 2021 from different locations in France. This dataset represents a valuable resource for studying and improving image processing techniques for astronomical photography on mobile devices. By providing a diverse and representative dataset of the current capabilities of these devices, we pave the way for new advancements in Deep Learning applied to mobile astrophotography. We hope AstroSmartphoneDataset will encourage the scientific community to explore innovative approaches for other Computer Vision tasks, object recognition, auto-labelling in images captured with smartphones. Thus, we have proposed a use-case consisting in labelling the images to train YOLO models for streaks detection, enabling to automatically track transient streak-like objects such as satellites or meteors across large datasets.

In future work, we will continue to capture additional images with other types of smartphones and from other locations, to to improve dataset generalizability. Moreover, we will use the data to design and train tiny Deep Learning models, with a view to embed them directly in smartphones: this could enable real-time inference, allowing users to instantly analyze their astrophotography images without the need for powerful computing infrastructures.

ACKNOWLEDGEMENTS

This research was funded by the Luxembourg Institute of Science and Technology (LIST). Data storage, data processing and Deep Learning models training were conducted on the LIST Artificial Intelligence and Data Analytics platform (AIDA), with the assistance of Jean-François Merche and Raynald Jadoul.

DATA AVAILABILITY

AstroSmartphoneDataset is available as a ZIP archive on Zenodo (https://zenodo.org/records/14933725), and a dedicated Github repository has also been set to aggregate related content (https://github.com/oparisot/AstroSmartphoneDataset). Additional materials used to support the results of this paper are available from the corresponding author upon request.

REFERENCES

Arimatsu, K. (2025). The OASES project: exploring the outer Solar System through stellar occultation with amateur-class telescopes. *Philosophical Transactions A*, 383(2291):20240191.

Ashley, J. (2017). *Video Astronomy on the Go*. Springer International Publishing.

Bortle, J. E. (2001). The bortle dark-sky scale. *Sky and Telescope*, 161(126):5–6.

Drechsler, M., Strottner, X., Sainty, Y., Fesen, R. A., Kimeswenger, S., Shull, J. M., Falls, B., Vergnes, C., Martino, N., and Walker, S. (2023). Discovery of extensive [o iii] emission near m31. *Research Notes of the AAS*, 7(1):1.

Falkner, D. E. (2020). Astrophotography using a compact digital camera or smartphone camera. In *The Mythology of the Night Sky*, pages 269–294. Springer.

Fang, Z., Ignatov, A., Zamfir, E., and Timofte, R. (2023). Sqad: Automatic smartphone camera quality assessment and benchmarking. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 20532–20542.

Farnebäck, G. (2003). Two-frame motion estimation based on polynomial expansion. In *Image Analysis: 13th Scandinavian Conference, SCIA 2003 Halmstad, Sweden, June 29–July 2, 2003 Proceedings 13*, pages 363–370. Springer.

Fisher, A. and Hinckley, S. (2022). Smartphone astrophotography. In *Proceedings of the IUPAP International Conference on Physics Education 2022*, pages 84–84.

Huang, C., Zeng, Q., Xiong, F., and Xu, J. (2024). Space dynamic target tracking method based on five-frame difference and deepsort. *Scientific Reports*, 14(1):6020.

Jegham, N., Koh, C. Y., Abdelatti, M., and Hendawi, A. (2025). YOLO Evolution: A Comprehensive Benchmark and Architectural Review of YOLOv12, YOLO11, and Their Previous Versions.

Kang, C. H. and Kim, S. Y. (2023). Real-time object detection and segmentation technology: an analysis of the yolo algorithm. *JMST Advances*, 5(2):69–76.

Kim, J.-H. and Ju, Y.-G. (2014). Fabrication of a video telescope using a smartphone. *New Physics*, 64(3):307–312.

Lécureuil, P. (2024). *Photographier l'Univers avec un smartphone: Petit guide d'initiation à l'astrophotographie*. De Boeck Supérieur.

Lee, A. (2023). *A Journey to the Edge of the Solar System with an AI navigator*. PhD thesis, University of Victoria.

Malchenko, S. L. (2024). From smartphones to stargazing: the impact of mobile-enhanced learning on astronomy education. *Science Education Quarterly*, 1(1):1–7.

Muñoz-Gil, G., Dauphin, A., Beduini, F. A., and Sánchez de Miguel, A. (2022). Citizen science to assess light pollution with mobile phones. *Remote Sensing*, 14(19):4976.

Nir, G., Zackay, B., and Ofek, E. O. (2018). Optimal and efficient streak detection in astronomical images. *The Astronomical Journal*, 156(5):229.

Parisot, O. (2025). AstroSmartphoneDataset: a collection of astronomical images captured with smartphones.

Parisot, O., Bruneau, P., Hitzelberger, P., Krebs, G., and Destruel, C. (2022). Improving accessibility for deep sky observation. *ERCIM News*, 2022(130).

Parisot, O. and Jaziri, M. (2024). Impact of Satellites Streaks for Observational Astronomy: A Study on Data Captured During One Year from Luxembourg Greater Region. In *Proceedings of the 13th International Conference on Data Science, Technology and Applications - Volume 1: DATA*, pages 417–424. INSTICC, SciTePress.

Paula, M. E. and Micha, D. N. (2021). Low-cost astrophotography with a smartphone: Steam in action. *The Physics Teacher*, 59(6):446–449.

Petit, J.-M., Holman, M., Scholl, H., Kavelaars, J., and Gladman, B. (2004). A highly automated moving object detection package. *Monthly Notices of the Royal Astronomical Society*, 347(2):471–480.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. (2021). Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR.

Romanov, F. (2022). The contribution of the modern amateur astronomer to the science of astronomy. *arXiv preprint arXiv:2212.12543*.

Sohib, A., Danusaid, N., Sawitri, A., Nuryadin, B. W., and Agustina, R. D. (2018). Low cost digitalization of observation telescope by utilizing smartphone. In *MATEC Web of Conferences*, volume 197, page 02002. EDP Sciences.

Wang, J., Chan, K. C., and Loy, C. C. (2023). Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 2555–2563.

Woodhouse, C. (2015). *The Astrophotography Manual*. Routledge.